

VU Research Portal

Development and validation of a HEXACO situational judgment test

Oostrom, Janneke K.; de Vries, Reinout E.; de Wit, Mariska

published in

Human Performance
2019

DOI (link to publisher)

[10.1080/08959285.2018.1539856](https://doi.org/10.1080/08959285.2018.1539856)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Oostrom, J. K., de Vries, R. E., & de Wit, M. (2019). Development and validation of a HEXACO situational judgment test. *Human Performance*, 32(1), 1-29. <https://doi.org/10.1080/08959285.2018.1539856>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



Development and validation of a HEXACO situational judgment test

Janneke K. Oostrom, Reinout E. de Vries & Mariska de Wit

To cite this article: Janneke K. Oostrom, Reinout E. de Vries & Mariska de Wit (2019) Development and validation of a HEXACO situational judgment test, Human Performance, 32:1, 1-29, DOI: [10.1080/08959285.2018.1539856](https://doi.org/10.1080/08959285.2018.1539856)

To link to this article: <https://doi.org/10.1080/08959285.2018.1539856>



Published online: 14 Nov 2018.



Submit your article to this journal [↗](#)



Article views: 792



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)



Development and validation of a HEXACO situational judgment test

Janneke K. Oostrom^a, Reinout E. de Vries^{a,b}, and Mariska de Wit^a

^aVrije Universiteit Amsterdam; ^bUniversity of Twente

ABSTRACT

The purpose of this study was to develop and validate a construct-based situational judgment test of the HEXACO personality dimensions. In four studies, among applicants, employees, and Amazon Mechanical Turk participants ($N_s = 72\text{--}305$), we showed that it is possible to assess the six personality dimensions with a situational judgment test and that the criterion-related validity of the situational judgment test is comparable to the criterion-related validity of traditional self-reports but lower than the criterion-related validity of other-reports of personality. Test-retest coefficients (with a time interval of 2 weeks) varied between .55 and .74. Considering personality is the most commonly assessed construct in employee selection contexts (Ryan et al., 2015), this situational judgment test may provide human resources professionals with an alternative assessment tool.

The importance of personality traits in predicting workplace behaviors has been well established (Sackett & Walmsley, 2014; Schmitt, 2014). Therefore, human resources (HR) professionals often rely on personality assessments. In fact, in a recent survey among HR professionals around the globe, Ryan et al. (2015) showed that personality is the most commonly assessed construct in employee selection contexts. Personality is often assessed through self-report measures. However, in the vast body of research on personality measures, there is a continuing debate on the value of self-report personality measures as they are susceptible to faking (Anglim, Morse, De Vries, MacCann, & Marty, 2017; Birkeland, Manson, Kisamore, Brannick, & Smith, 2006), rely on the ability and motivation to introspect accurately (De Cuyper et al., 2017) and can be influenced by a variety of biases and response sets (e.g., consistency motivation; Paulhus & Vazire, 2007). Considering the ubiquity of self-report personality measures in employee selection contexts and the disagreement surrounding their use, a continued search for alternative techniques to measure personality is warranted.

When searching for alternative techniques to measure personality, it is important to make a distinction between predictor constructs and predictor methods (Arthur & Villado, 2008). Predictor constructs refer to the behavioral domain being sampled, and predictor methods refer to the technique by which domain-relevant responses are elicited. In the context of personality measures, research has revealed that the criterion-related validity of self-reports can be increased by a change in the technique by which responses are elicited, that is, by means of contextualization (e.g., Lievens, De Corte, & Schollaert, 2008; Schaffer & Postlethwaite, 2012). Contextualized personality measures that specifically refer to the workplace instruct candidates to describe themselves exclusively in terms of their workplace behaviors, add the label “at work” to each item, or fully contextualize the items (Morgeson et al., 2007). Fully contextualized personality measures, which contain completely redesigned items to match a specific context generally show the largest increases in criterion-related validity (e.g., Holtrop, Born, De Vries, & De Vries, 2014).

A situational judgment test (SJT) is an example of a fully contextualized measure. However, compared with most fully contextualized measures (e.g., Holtrop et al., 2014), which typically use a Likert-type response scale, a typical SJT presents candidates with work situations followed by several possible responses (e.g., Oostrom, De Soete, & Lievens, 2015). There is ample evidence for the criterion-related validity of SJTs (e.g., Christian, Edwards, & Bradley, 2010; Oostrom, Born, Serlie, & Van der Molen, 2012). Furthermore, SJTs show low levels of adverse impact (Whetzel, McDaniel, & Nguyen, 2008) and are generally accepted among candidates (Whetzel & McDaniel, 2009). There is also evidence suggesting that compared to self-report personality measures, SJTs are less susceptible to faking and are less dependent upon a candidate's ability to use introspection (Hooper, Cullen, & Sackett, 2006; Mussel, Gatzka, & Hewig, 2016).¹ So far, only a few SJTs have been developed to measure personality (e.g., Motowidlo, Hooper, & Jackson, 2006; Mussel et al., 2016). However, as far as we know, there is no published SJT of the HEXACO personality model. The goal of the present study therefore is to develop and validate an SJT of the six HEXACO personality dimensions.

Personality assessment

Most personality measures that are used by practitioners are based on the Five-Factor Model or the Big Five factors, generally known as Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience (e.g., Goldberg, 1990; Norman, 1963; Tupes & Christal, 1961). Although these personality factors are widely used by researchers and practitioners, a reanalysis of the lexical data that has become available from at least a dozen languages, including English, has recovered a maximum set of six cross-culturally replicable personality factors instead of five (e.g., Ashton, Lee, & De Vries, 2014; Ashton et al., 2006; Ashton, Lee, & Goldberg, 2004; De Raad et al., 2014; Saucier, 2009). The results of these lexical investigations revealed not only variants of the Big Five factors (i.e., Emotionality and Agreeableness correspond roughly to rotated variants of the Big Five Emotional Stability and Agreeableness axes) but also a sixth factor named Honesty-Humility. In recent years, considerable evidence has accumulated in favor of this alternative representation of the personality structure, known as the HEXACO model (e.g., Ashton & Lee, 2007). In 2000, Lee and Ashton began the construction of a new personality measure based on the HEXACO model (Lee & Ashton, 2004). The psychometric properties of this inventory (the HEXACO-PI and its revision, the HEXACO-PI-R) have been well-established by now (e.g., De Vries, Ashton, & Lee, 2009; Lee & Ashton, 2004).

Yet there is a disadvantage to using self-report measures, including the HEXACO-PI-R, for assessing personality in employee selection contexts. As just described, the criterion-related validity of self-report measures may be less than optimal because they assess global, noncontextualized personality factors (e.g., Pomerance & Converse, 2014). There is ample research indicating that contextualization benefits the criterion-related validity of personality measures (e.g., Eschleman & Burns, 2012; Pomerance & Converse, 2014; Shaffer & Postlethwaite, 2012). For example, Shaffer and Postlethwaite (2012) found that the validities for noncontextualized personality measures range from .02 to .22 (with a mean of .11), whereas the validities for contextualized personality measures range from .14 to .30 (with a mean of .24). For three of the five personality scales, they found that the highest level of contextualization (i.e., contextualizing items and instructions) leads to the highest criterion-related validity. In addition, Holtrop et al. (2014) pointed out that the face validity and perceived predictive validity of the HEXACO-PI-R improves by contextualizing its items.

The explanation for why contextualized personality measures have greater predictive validity than general personality measures pertains to the frame-of-reference effect (Schmit, Ryan, Stierwalt, & Powell, 1995). The theoretical basis for exploring frame-of-reference effects on the validity of personality measures can be found in person-situation interaction and trait activation theories (De Vries, Tybur, Pollet, & Van Vugt, 2016; Lievens, Chasteen, Day, & Christiansen, 2006; Mischel, 1973; Tett & Burnett, 2003), which rest on the notion that personality is not necessarily a consistent predictor of behavior across situations. Specifically, the person-situation interaction and trait

activation theories predict that behavior is a function of both the personality of the individual and the situation itself: Some situations may be powerful determinants of behavior, whereas other situations are not or only trivially so (De Vries, Tybur et al., 2016; Lievens et al., 2006; Mischel, 1973; Tett & Burnett, 2003). Consequently, these conditional models of dispositions predict that some individuals may be more likely to demonstrate certain behaviors at work but not at home, or vice versa, depending on the extent to which traits are differentially activated in work and home situations and the extent to which incentives are involved for exhibiting certain behaviors in these situations. Because traits are differentially activated and rewarded in different situations, individuals are likely to differ somewhat in trait expression across situations. By contextualizing a personality measure, this differential expression of traits is taken into account. Contextualization of a personality measure leads to the use of a specific frame-of-reference (Lievens et al., 2008). Because of this frame-of-reference, every respondent will have the same situational context in mind when completing each of the items, which not only ensures a better alignment of the responses to the actual exhibited behavior in the situation but also increases the conceptual similarity of the personality measure and the behavior that it aims to predict.

Situational judgment tests

Over the past decade, SJTs have become increasingly popular in research and in practice (Weekley & Ployhart, 2005). An SJT is a measurement method typically composed of job-related situations that describe a work situation, followed by a list of plausible response options. Candidates are asked to evaluate each response option for either the likelihood that they would respond in that particular manner or the effectiveness of the response. In general, the literature has supported the predictive validity of SJTs (e.g., Christian et al., 2010; Oostrom et al., 2012). For instance, in a meta-analysis, McDaniel, Hartman, Whetzel, and Grubb (2007) found SJT scores to have an average observed validity of .20 for predicting job performance. Clevenger, Pereira, Wiechmann, Schmitt, and Harvey (2001) showed that SJTs are able to explain incremental variance in job performance when controlling for the effects of other more traditional predictors, such as cognitive ability, personality, and job experience.

Most SJTs are developed based on a work-sample-based approach in which critical incidents are identified either from archival records or from interviews with subject matter experts such as managers, employees, or other key stakeholders (Weekley, Ployhart, & Holtz, 2006). These SJTs are developed for very specific contexts (i.e., a specific occupation at a specific organization) and therefore have the highest level of contextualization. However, these SJTs are not explicitly designed to measure any particular psychological construct. In fact, it often remains unclear what constructs are associated with behaviors sampled by these SJTs (McDaniel, List, & Kepes, 2016). It is therefore unlikely that this traditional development approach would allow for the development of an SJT that measures the six HEXACO dimensions.

Recently, Lievens (2017) described an alternative way of developing and scoring SJTs to assess specific constructs, which may provide higher levels of construct-related validity. There are two important differences between the traditional work-sample-based approach and this alternative construct-driven approach to SJT development. First, response options of construct-driven SJTs are more unidimensional because they lie on a continuum with each response option reflecting a different level of the trait. This allows for the computation of a trait score in the same way as in a personality measure, that is, by averaging a candidate's endorsement. Second, the items of construct-driven SJTs are not necessarily developed with the help of subject matter experts. Instead, psychologists are typically in charge of developing the items on the basis of trait activation theory (Tett & Burnett, 2003). Most construct-driven SJTs present work-related situations that generalize across occupations and organizations. Construct-driven SJTs are therefore less contextualized than traditional SJTs but still provide more context than most contextualized personality measures (e.g., Holtrop et al., 2014).

A construct-driven approach in the development of SJTs offers several advantages over the traditional work-sample-based approach. First, the specification of the construct domain reduces contamination due to the measurement of unintended constructs (Christian et al., 2010). Second, it helps to disentangle the effects of the measurement method from the construct, which helps to understand why the SJT is related to the criterion of interest (Arthur & Villado, 2008). Third, it offers the possibility to develop tests that will generalize across different occupations and industries (Lievens, 2017; Motowidlo, Ghosh, Mendoza, Buchanan, & Lerma, 2016). Fourth, it permits more accurate feedback (i.e., at the construct level) in development settings (Guenole, Chernyshenko, & Weekly, 2017). Despite these advantages, most SJTs are still based on the work-sample approach (Christian et al., 2010). However, the few SJTs that have been developed to assess specific constructs show promising results. For example, De Meijer, Born, Van Zielst, and Van der Molen (2010) developed an SJT to measure the construct of integrity, and Bledow and Frese (2009) developed an SJT to measure the construct of personal initiative. Both found support for the construct-related validity of their SJT.

Present study

The goal of the present study is to develop an SJT for the HEXACO personality dimensions and to examine its construct-related and criterion-related validity. In Study 1, we describe the development of the SJT and test its convergent and discriminant validity by relating its scores to self-reported HEXACO scores. In Study 2, we test the factor structure and convergent and discriminant validity of the SJT in an applicant sample and relate its scores to the selection outcome. In Study 3, we again test the factor structure and the convergent and discriminant validity of the SJT by relating its scores to both self-reported and other-reported HEXACO scores and we test its concurrent validity for predicting task performance and organizational citizenship behavior (OCB). In Study 4, we examine the test-retest reliability of the SJT and compare the test-retest reliability, factor structure, and test difficulty (i.e., content difficulty and perceived difficulty) of two versions of the SJT: one in which the response options are presented on a continuum (either from low to high or from high to low) and one in which the SJT response options are presented in a randomized order.

Study 1 method

Sample and procedure

Participants were recruited through social media; we posted an invitation to participate in our online survey on LinkedIn and Facebook. The total data set consisted of 72 participants (31 men), with a minimum age of 16 and a maximum age of 62 years old ($M_{\text{age}} = 33.06$, $SD = 13.88$). This sample size gave us a power of 74% to detect medium effects ($r = .30$; Cohen, 1992). Participants could take part in this study only if they worked for at least 2 days a week. The participants had an average of 13.56 years ($SD = 12.68$) of work experience. They received a link to a survey containing the HEXACO-SJT and the 60-item HEXACO-PI-R (i.e., HEXACO-60).

Measures

SJT

Constructing the SJT was an iterative process, in which current recommended practice for construct-driven SJT development was followed (e.g., Guenole et al., 2017; Lievens, 2017). A behavioral tendency instruction was used (i.e., “What would you do?”), as this type of instruction is more suitable to measure personality than a knowledge-based instruction (Whetzel & McDaniel, 2009). A total of 182 trained undergraduate psychology students (35 men, $M_{\text{age}} = 23.18$, $SD = 2.19$) helped develop the SJT items as part of their Measurement and Diagnostics class. In line with trait activation

theory (Tett & Burnett, 2003), they were instructed to develop SJT item stems that would elicit behaviors that were indicative of a targeted HEXACO trait. The response options had to express different levels of the trait. They developed items in groups of two to four, after which the first and second author of this article made a selection of four items per HEXACO dimension. The students then completed the SJT and the 100-item HEXACO measure (Ashton & Lee, 2008; De Vries et al., 2009). As suggested by Guenole et al. (2017), multitrait multimethod (MTMM) methods were used to assess convergent and discriminant validities. Only items that showed the highest correlation with the intended trait were retained; the other items were rewritten or replaced by new items. This process was repeated several times, among three samples of students (total sample sizes varied between 54 and 70), a sample of actual applicants ($N = 101$), and a sample of employees recruited via social media ($N = 67$).

The final SJT consisted of 24 items (four items per HEXACO dimension) and took about 20–30 min to complete. Each response option reflected a different level of the trait, with 1 at the lowest level of the trait and 4 at the highest level of the trait. Participants' SJT scores were represented by their average trait score on the four items.² The final SJT items are included in the appendix. Cronbach's alphas are included in the correlation tables. The alphas were expected to be low because the four items within each dimension measure a mix of underlying facets. Note that the alphas were comparable to the alphas of the brief version of the HEXACO-PI-R (De Vries, 2013), which also consists of four items per dimension.

Hexaco

We used the HEXACO-60 (Ashton & Lee, 2009; De Vries et al., 2009) to measure the HEXACO personality dimensions. Participants were asked for their agreement (1 = *strongly disagree*; 5 = *strongly agree*) with the statements. An example statement measuring Conscientiousness is "I plan ahead and organize things, to avoid scrambling at the last minute." The items were averaged to create a total score of the six dimensions. Cronbach's alpha was .77 for Honesty-Humility, .75 for Emotionality, .79 for Extraversion, .71 for Agreeableness, .71 for Conscientiousness, and .77 for Openness to Experience.

Study 1 results

Table 1 presents the means, standard deviations, and intercorrelations for the HEXACO dimensions as measured with the SJT and the HEXACO-60. Results showed support for the construct-related validity of the SJT. The convergent correlations ranged from .34 ($p < .01$) for Agreeableness to .78 ($p < .01$) for Openness to Experience. The discriminant correlations ranged from $-.30$ ($p < .05$) between Agreeableness as measured with the SJT and Extraversion as measured with the HEXACO-60 to .39 ($p < .01$) between Extraversion as measured with the SJT and Openness to Experience as measured with the HEXACO-60.

Table 2 presents the correlations between the SJT scales and the HEXACO scales measuring the same construct (monotrait-heteromethod correlation) and the average correlation between the SJT scale and the HEXACO scales measuring different constructs (heterotrait-heteromethod correlation) in Studies 1 to 3. Steiger's (1980) z statistic was used to calculate the difference between the monotrait-heteromethod and each heterotrait-heteromethod correlation. In Table 2 we report the mean z values. The results of Study 1 show that, for each construct, the convergent correlation was higher than the average discriminant correlation. Steiger's z showed a significant difference (all $ps < .01$) between the two correlation coefficients for all constructs.

In addition, we used a confirmatory factor analysis (CFA) to evaluate the convergent and discriminant validities (Marsh & Grayson, 1995). We examined the correlated trait-correlated uniqueness (CTCU) model, as this model is preferred over the correlated trait-correlated method model when multidimensional method effects are present (Tomas & Oliver, 1999). Factor loadings within each trait were constrained to be equal. We first tested a starting model specifying correlated trait factors but no method factors. This model yielded a relatively poor fit to the data, χ^2

Table 1. Means, standard deviations, and intercorrelations for all variables in Study 1

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12
1. SJT-H	3.40	0.46	(.50)											
2. SJT-E	2.14	0.42	-.07	(.22)										
3. SJT-X	2.87	0.45	-.13	-.09	(.50)									
4. SJT-A	2.24	0.42	-.10	.17	.03	(.26)								
5. SJT-C	2.89	0.51	.22	.06	-.21	.12	(.38)							
6. SJT-O	2.91	0.51	.28*	-.05	.40*	-.03	.02	(.66)						
7. Honesty-Humility	3.51	0.51	.50**	-.18	-.02	.04	.16	.21	(.77)					
8. Emotionality	2.83	0.62	.03	.52**	-.14	.08	.17	.07	-.11	(.75)				
9. Extraversion	3.58	0.50	.08	-.19	.53**	-.30*	-.05	.23	.18	-.33	(.79)			
10. Agreeableness	3.12	0.58	.15	-.06	-.05	.34**	.10	.13	.47**	-.09	-.14	(.71)		
11. Conscientiousness	3.55	0.57	.23	-.13	-.15	-.04	.55**	.07	.15*	.27*	-.23	.08	(.71)	
12. Openness to Experience	3.13	0.62	.26*	-.06	.39**	-.04	-.08	.78**	.24*	.07	.21	.02	-.02	(.77)

Note. *N* = 72. The situational judgment test (SJT) HEXACO dimensions were measured on a scale from 1 to 4, and the HEXACO dimensions as measured with the HEXACO-PI-R were measured on a scale from 1 to 5. Cronbach's alphas are presented on the diagonal between parentheses. Significant convergent correlations are presented in boldface.

p* < .05 (two-tailed). *p* < .01.

Table 2. Overview of monotrait-heteromethod and heterotrait-heteromethod correlations in all studies

	Study 1			Study 2 ^a			Study 3 Self-Ratings			Study 3 Other-Ratings		
	Mono-trait	Hetero-trait	<i>z</i>	Mono-trait	Hetero-trait	<i>z</i>	Mono-trait	Hetero-trait	<i>z</i>	Mono-trait	Hetero-trait	<i>z</i>
Honesty-Humility	.50	.15	2.53**	.32	.06	2.41**	.31	.11	1.61	.22	.09	0.97
Emotionality	.52	-.13	4.11**	.10	.09	0.10	.58	.00	4.64**	.39	.10	2.07**
Extraversion	.53	.01	3.24**	.44	.11	3.19**	.56	.06	4.11**	.36	.07	2.17**
Agreeableness	.34	-.05	2.47**	.15	.07	1.04	.30	.09	1.63	-.01	.05	-.035
Conscientiousness	.55	.06	3.36**	.41	.10	2.96**	.42	.06	2.93**	.26	.06	1.53
Openness to Experience	.78	.14	5.49**	.53	.11	4.22**	.64	-.04	5.81**	.39	.15	1.85*

Note. The monotrait-heteromethod correlation is the correlation between the situational judgment test (SJT) scale and the HEXACO scale measuring the same construct. The heterotrait-heteromethod correlation is the average of the correlations between the SJT scale and the HEXACO scales measuring different constructs. We used the Fisher *r*-to-*z* transformation before averaging the correlations and then changed the mean *z* value back to *r* using the same transformation. Steiger's (1980) *z* statistic was used to calculate the difference between the monotrait-heteromethod correlation and all heterotrait-heteromethod correlations. In the table we present the mean *z* values.

^aIn Study 2, Emotionality (Neuroticism), Extraversion, Agreeableness, Conscientiousness, and Openness to Experience were measured with the NEO-PI-R.

p* < .05 (one-tailed). *p* < .01.

(45) = 87.48, *p* < .01, comparative fit index (CFI) = .81, root mean square error of approximation (RMSEA) = .12, standardized root mean square residual (SRMR) = .09. Next, the CTCU model was evaluated. This model yielded a better fit to the data, $\Delta\chi^2(30) = 69.48$, *p* < .01; $\chi^2(15) = 18.00$, *p* = .26; CFI = .99, RMSEA = .05, SRMR = .06. The standardized parameter estimates of the CTCU model are presented in Table 3. Consistent with the MTMM correlations, convergent validities across the two measures were good; all trait factor loadings were substantial and significant, indicating that the two measures generally agreed in their estimation of the personality trait constructs. Only four intercorrelations among the latent HEXACO factors were significant, indicating adequate discriminant validity. As indicated by the uniqueness correlations (Kenny & Kashy, 1992), common method variance was present in both the SJT (*M* absolute *r* = .34) and the HEXACO-60 (*M* absolute *r* = .24).

Study 1 discussion and study 2 introduction

The results of this first study show some initial evidence that our HEXACO-SJT is measuring the intended personality dimensions. The MTMM correlations showed that five of the six convergent

Table 3. Correlated traits–correlated uniqueness standardized parameter estimates in Study 1

	H	E	X	A	C	O
Factor loadings						
SJT	.72**	.91**	.82**	.67**	.75**	.95**
HEXACO	.64**	.58**	.72**	.48**	.70**	.81**
Latent factor correlations						
H	1.00					
E	−.18	1.00				
X	.07	−.29*	1.00			
A	.34	.00	−.37*	1.00		
C	.32	−.03	−.09	.10	1.00	
O	.39**	−.06	.51**	.04	−.01	1.00
Uniqueness correlations						
SJT						
H	1.00					
E	.23	1.00				
X	−.46*	.51	1.00			
A	−.39	.41	.66**	1.00		
C	.10	.22	−.18	.29	1.00	
O	−.11	−.04	.06	−.07	.29	1.00
HEXACO						
H	1.00					
E	−.12	1.00				
X	.30	−.37*	1.00			
A	.50**	−.10	−.09	1.00		
C	−.08	.42**	−.39*	.00	1.00	
O	.14	.08	−.10	−.13	−.10	1.00

Note. SJT = situational judgment test.

* $p < .05$. ** $p < .01$.

validities were above .50, whereas the mean convergent validity among personality scales is generally below .50 (Pace & Brannick, 2010). In addition, the CTCU model showed that the SJT and the HEXACO-60 agreed in their estimation of the personality trait constructs. However, the sample size of Study 1 is relatively small and did not allow a test of the factor structure of the HEXACO-SJT by means of a CFA. Power analysis shows that we would need at least 105 participants to have sufficient power to test the factor structure of our SJT (with $df = 237$; MacCallum, Browne, & Sugawara, 1996). Because the factor structure of both personality measures and SJTs is known to be different in employee samples compared to applicant samples (e.g., MacKenzie, Ployhart, Weekley, & Ehlers, 2009; Schmit & Ryan, 1993), it is important to test the validity of the SJT using an applicant sample. Hence, the goal of Study 2 is to test the factor structure of the SJT and its validity for predicting the selection outcome (on which the SJT scores had no bearing) in an applicant sample.

Study 2 method

Sample and procedure

Participants were Dutch military applicants. In a period of 4 weeks, all candidates ($N = 478$) who applied for the military received an e-mail in which they were asked to complete additional measures (i.e., the HEXACO-SJT and the Honesty-Humility scale) 8 days before the actual selection procedure. The actual selection procedure consists of a personality measure, several capacity tests, an achievement motivation measure, and a 60- to 90-min life span interview. Removing cases that were incomplete, partially random, or inattentive, yielded a complete data set of $N = 157$ (125 men, $M_{\text{age}} = 22.27$, $SD = 5.26$). This sample size gave us a power of 97% to detect medium effects ($r = .30$; Cohen, 1992). Study participation was voluntary, and participants were explained that participation would not influence the outcomes of the selection procedure.

Measures

SJT

We used the same SJT as in Study 1.

Personality measures

The Ministry of Defense uses the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) to measure applicants' Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each scale is measured using 48 items that ask respondents to indicate the accuracy of different statements about themselves with response options ranging from 1 (*very inaccurate*) to 5 (*very accurate*). We were only able to retrieve stanine scores, ranging from 1 (*very low*) to 9 (*very high*) based on general selection norms. The test manual reported adequate Cronbach's alphas for the five scales (Neuroticism = .92, Extraversion = .89, Openness to Experience = .87, Agreeableness = .86, Conscientiousness = .90) and test-retest coefficients after 6 years (Neuroticism = .83, Extraversion = .82, Openness to Experience = .83, Agreeableness = .63, Conscientiousness = .79).

Participants also completed the 10 Honesty-Humility items of the HEXACO-60 (Ashton & Lee, 2009; De Vries et al., 2009). An example item is "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed." Cronbach's alpha was .62.

Selection outcome

The assessors conducted a 60-min to 90-min semistructured life span interview with each applicant. Directly after the interview, the assessors received the applicant's scores on the personality measure, capacity test, and achievement-motivation measure. After reviewing all scores, the assessors decided on the selection outcome, which was either 0 (*reject*) or 1 (*continue to the next selection round*).

Study 2 results

Table 4 presents the means, standard deviations, and intercorrelations of the HEXACO-SJT, the Five Factor Model personality traits, and Honesty-Humility. Results showed support for the construct-related validity for four of the SJT scales—Honesty-Humility, Extraversion, Conscientiousness, and Openness to Experience. For these scales, the convergent correlations ranged from .32 ($p < .01$) for Honesty-Humility to .53 ($p < .01$) for Openness to Experience. The discriminant correlations ranged from $-.34$ ($p < .01$) between Openness to Experience as measured with the SJT and Neuroticism as measured with the NEO-PI-R and .37 ($p < .01$) between Openness to Experience as measured with the SJT and Conscientiousness as measured with the NEO-PI-R. Table 2 (Study 2) shows that for all constructs the convergent correlation was significantly higher than the average discriminant correlation, except for the two constructs that are known to differ between the HEXACO and the NEO models, that is, HEXACO Emotionality (vs. NEO Neuroticism) and HEXACO (vs. NEO) Agreeableness.

We also used a CFA to evaluate the convergent and discriminant validities. As the covariance matrices were not positively defined, neither the trait only nor the CTCU model yielded an admissible solution. The covariance matrices showed negative covariances for Emotionality/Neuroticism and Agreeableness. We therefore decided to remove these two scales from the analyses. The starting model, specifying correlated trait factors but no method factors, yielded a good fit to the data, $\chi^2(18) = 29.58$, $p = .04$, CFI = .95, RMSEA = .06, SRMR = .06. The CTCU model yielded a slightly but not significantly better fit to the data, $\Delta\chi^2(12) = 16.28$, $p = .18$; $\chi^2(6) = 13.30$, $p = .04$; CFI = .97, RMSEA = .09, SRMR = .04. The standardized parameter estimates of the CTCU model are presented in Table 5. All trait factor loadings were substantial and significant, indicating that the measures agreed in their estimation of the personality trait constructs. The intercorrelations among the latent HEXACO factors showed poor discriminant validity. This is likely due to the high-stakes setting in which the NEO-PI-R was administered, causing

Table 4. Means, standard deviations, and intercorrelations for all variables in Study 2

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
1. SJT-H	3.55	0.27	(.34)												
2. SJT-E	1.93	0.40	.04	(.39)											
3. SJT-X	2.92	0.38	.11	.02	(.44)										
4. SJT-A	2.55	0.48	.16	-.16*	.08	(.52)									
5. SJT-C	3.18	0.49	.11	.08	.31**	.16*	(.50)								
6. SJT-O	2.83	0.54	.24**	.06	.37**	.12	.25**	(.66)							
7. Honesty-Humility	3.85	0.45	.32**	.02	.15	.21**	.21**	.20*	(.62)						
8. Neuroticism	3.67	1.89	-.14	.10	-.32**	-.03	-.25**	.17*	-.29**	(.92)					
9. Extraversion	5.00	1.71	-.02	.13	.44**	.02	.26**	.15	.03	-.37**	(.89)				
10. Agreeableness	5.12	1.69	.23**	.19*	.17*	.15	.19*	.15	.29**	-.33**	.24**	(.86)			
11. Conscientiousness	5.27	1.93	.13	-.02	.34**	.09	.41**	.37**	.27**	-.67**	.41**	.32**	(.90)		
12. Openness to Experience	3.72	1.79	.11	.11	.18*	-.07	.06	.53**	.13	-.15	.19*	.05	.26**	(.87)	
13. Selection Outcome	0.73	0.45	.04	-.02	.12	-.02	.23**	.01	.05	-.26*	.20*	.06	.19*	.06	(-)

Note. *N* = 157. The situational judgment test (SJT) HEXACO dimensions were measured on a scale from 1 to 4, Honesty-Humility was measured on a scale from 1 to 5, the Big Five personality dimensions as measured with the NEO-PI-R were measured on a scale from 1 to 9, and the selection outcome was coded as 0 (*reject*) and 1 (*continue to the next round*). Cronbach's alphas are presented on the diagonal between parentheses. Significant convergent correlations are presented in boldface.

p* < .05 (two-tailed). *p* < .01.

Table 5. Correlated traits–correlated uniqueness standardized parameter estimates in Study 2

	H	E/N	X	A	C	O
Factor loadings						
SJT	.56**	–	.67**	–	.62**	.74**
NEO-PI-R and Honesty-Humility scale	.56**	–	.66**	–	.64**	.72**
Latent factor correlations						
H	1.00					
E	–	—				
X	.16	–	1.00			
A	–	–	–	–		
C	.47**	–	.69**	–	1.00	
O	.34*	–	.36**	–	.51**	1.00
Uniqueness correlations						
SJT						
H	1.00					
E	–	–				
X	–.07	–	1.00	–		
A	–	–	–	–		
C	.11	–	.21	–	1.00	
O	–.03	–	.09	—	.01	1.00
HEXACO						
H	1.00					
E	–	–				
X	.09	–	1.00			
A	–	–	–	–		
C	–.03	–	.03	–	1.00	
O	.17	–	.34*	–	.09	1.00

Note. As the Honesty-Humility scale and the Revised NEO Personality Inventory (NEO-PI-R) were measured on different scales, we used the standardized scores in this analysis. SJT = situational judgment test.

* $p < .05$. ** $p < .01$.

higher intercorrelations among its scales (see also Table 4). As indicated by the uniqueness correlations, common method variance was lower in the SJT (mean absolute $r = .09$) than in the NEO-PI-R/Honesty-Humility scale (mean absolute $r = .13$).

In addition, we assessed the factor structure of the SJT with a CFA. The CFA confirmed that a 6-factor model fit the data reasonably well, $\chi^2(237) = 311.22$, $p < .01$, CFI = .80, RMSEA = .05, SRMR = .07. Item loadings varied between .02 and .81. Most items ($k = 20$) had an item loading $> .30$. The six-factor model fit the data significantly better than a one-factor model, $\Delta\chi^2(15) = 111.32$, $p < .01$, $\chi^2(252) = 422.54$, $p < .01$, CFI = .54, RMSEA = .07, SRMR = .09.

The Conscientiousness scale of the SJT significantly correlated with the selection outcome ($r = .23$, $p < .01$). Of the NEO-PI-R scales, Neuroticism ($r = -.26$, $p < .01$), Extraversion ($r = .20$, $p < .05$), and Conscientiousness ($r = .19$, $p < .05$) significantly correlated with the outcome measure. To examine which scales explained most of the variance in the binary selection outcome, we conducted a logistic regression analysis with the NEO-PI-R and the Honesty-Humility scales in Step 1 and the SJT scales in Step 2. As per Table 6, the Neuroticism scale of the NEO-PI-R ($B = -0.37$, Wald = 5.92, $p < .05$) and the Conscientiousness scale of the SJT ($B = 1.13$, Wald = 5.59, $p < .05$) turned out to be the best predictors of the selection outcome.

Study 2 discussion and study 3 introduction

The results of Study 2 show further evidence for the construct-related validity of our HEXACO-SJT. First of all, the convergent and discriminant correlations showed a satisfactory pattern for four of the SJT scales, that is Honesty-Humility, Extraversion, Conscientiousness, and Openness to Experience. The two SJT scales that did not show a satisfactory pattern are Emotionality and Agreeableness. These HEXACO traits have been shown to roughly correspond to rotated variants of “Big Five” Emotional Stability and Agreeableness (e.g., Ashton & Lee, 2007). The low convergent correlations

Table 6. Results of stepwise logistic regression analyses in Study 2

	Step 1				Step 2			
	<i>B</i>	<i>SE (B)</i>	<i>Wald</i>	<i>Exp (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>Wald</i>	<i>Exp (B)</i>
Step 1								
Honesty-Humility	−0.05	0.46	−0.01	0.95	−0.26	0.49	0.28	0.77
Neuroticism	−0.30	0.14	4.46*	0.74	−0.37	0.15	5.92*	0.69
Extraversion	0.18	0.13	1.98	1.19	0.13	0.14	0.81	1.14
Agreeableness	−0.06	0.13	0.23	0.94	−0.06	0.14	0.18	0.94
Conscientiousness	0.02	0.14	0.01	1.02	−0.06	0.15	0.19	0.94
Openness to Experience	0.00	0.11	0.00	1.00	0.13	0.14	0.86	1.13
Step 2								
SJT-H					0.48	0.72	0.44	1.61
SJT-E					−0.17	0.53	0.10	0.84
SJT-X					0.11	0.64	0.03	1.12
SJT-A					−0.12	0.44	0.08	0.89
SJT-C					1.13	0.48	5.59*	3.08
SJT-O					−0.86	0.50	2.95	0.42
Nagelkerke R^2	.12				.19			

Note. $N = 157$. Dependent variable is the selection outcome, which was coded as 0 (*reject*) and 1 (*continue to the next round*). SJT = situational judgment test.

* $p < .05$. ** $p < .01$.

for Emotionality and Agreeableness seem to be in line with the findings of Lee and Ashton (2013) who also found low convergence between Neuroticism and Agreeableness as measured with the NEO-FFI (the 60-item version of the NEO-PI-R) and Emotionality and Agreeableness as measured with the HEXACO (in observer reports). Second, for most SJTs and personality measures it has been hard to find a good model fit (e.g., Hopwood & Donnellan, 2010; McDaniel & Whetzel, 2005), especially in applicant settings (e.g., MacKenzie et al., 2009; Schmit & Ryan, 1993). In fact, verification of factorial validity is rarely provided in SJT research (Kasten & Freund, 2015). Hence, the CFA results provide valuable insights into the factor structure of our HEXACO-SJT. Third, of the SJT scales, only the Conscientiousness scale correlated with the selection outcome. Note that this correlation was stronger than the correlation of four of the five NEO-PI-R scales which – in contrast to the SJT scales – were part of the selection procedure and were used by the assessors in their final selection decision. We have to note that we informed the applicants that their SJT performance would not influence the outcomes of the selection procedure. Therefore, it is important to replicate these findings in a high-stakes setting, in which applicants may have a stronger motivation to respond in a socially desirable way.

In sum, Studies 1 and 2 showed promising results regarding the construct-related validity of our HEXACO-SJT based on the correlations with self-report personality measures. Because other-reports of personality offer a better explanation and prediction of criteria (Connelly & Ones, 2010; De Vries, 2012; De Vries, Lee, & Ashton, 2008; Oh, Wang, & Mount, 2011), it is important to correlate the SJT scales to other-reports of personality. Hence, the first goal of Study 3 is to again test the convergent and discriminant validity of the SJT by relating its scores to both self-reported and other-reported HEXACO dimensions.

Furthermore, to be able to use the SJT as a predictor of job behaviors, it has to be correlated with actual job behaviors. Therefore, the second goal of Study 3 is to test the (incremental) validity of the SJT for predicting both task performance and OCB. The advantage of a construct-based SJT is that it allows specific predictions regarding its relations with external criteria (Arthur & Villado, 2008). Considering that personality dimensions generally show stronger relations with OCB than with task performance (e.g., Borman & Motowidlo, 1997; Chiaburu, Oh, Berry, Li, & Gardner, 2011), we consider OCB rather than task performance to be the relevant outcome measure for determining the criterion-related validity of the SJT. In our specific sample of accountants, we believe that predictable and responsible behaviors (Chiaburu et al., 2011) as well as honesty and modesty (Day & Silverman, 1989) are highly valued. Hence, we expect the Conscientiousness and Honesty-Humility scales of the SJT to be positively correlated with OCB.

Study 3 method

Sample and procedure

In collaboration with the HR department of a large accountancy firm in the Netherlands, an invitation was sent to all employees ($N = 143$) of a specific service line to participate in our study. In total, 110 employees (response rate = 76.92%) participated, which gave us a power of 89% to detect medium effects ($r = .30$; Cohen, 1992). Most participants were men ($n = 82$) and had a university degree (69.1%). On average, they were 34.52 years old ($SD = 10.18$) and had 12.02 years ($SD = 9.81$) of work experience. We mentioned in the invitation e-mail that all data would be treated confidentially, that participants would not see the scores of the other-report or the supervisor ratings, and that for every completed set of surveys (i.e., the self-report, the other-report, and the supervisor ratings) we would donate five euros to a good cause. The invitation e-mail also contained the link to the survey with the HEXACO-SJT and the HEXACO-PI-R items. At the end of the survey, participants were asked to provide us with a name and e-mail address of a colleague, friend, partner, or family member we could contact to obtain the other-report of the HEXACO dimensions. In addition, to measure the criterion variables, the HR department provided us with the list of names and e-mail addresses of the employees' managers, who were subsequently sent a link to a survey that contained questions about the employees' task performance and OCB. On 96 employees we obtained other-reports (47 men, $M_{\text{age}} = 37.24$, $SD = 11.34$), which were mostly from partners (42.7%), colleagues (21.8%), and family members (14.5%), and on 103 employees we obtained supervisor ratings (85 men, $M_{\text{age}} = 44.37$, $SD = 8.18$). In total, 40 managers provided ratings for 1–13 employees, of whom 22 managers provided ratings of two or more employees. The intraclass correlation (ICC) (1) of .36 and ICC(2) of .69 indicated a considerable amount of homogeneity of ratings within these 22 managers. However, a substantial number of managers rated only one employee, which resulted in a data set far from the minimum number of managers and employees that should be assessed by each manager to ensure sufficient power for multilevel analyses (e.g., Maas & Hox, 2005). Because the individual-level manager ratings are correlated with ratings from other sources (self-reports and other-reports of personality), we believe the manager ratings represent actual individual differences in task performance and OCB, and thus we decided to perform the analyses at the individual level.

Measures

SJT

We used the same SJT as in Studies 1 and 2.

Hexaco

We used the 208-item HEXACO-PI-R to measure self-reported HEXACO personality dimensions and the 104-item HEXACO-PI-R to measure other-reported personality dimensions (Ashton & Lee, 2009; De Vries et al., 2008).³ Participants were asked for their agreement, from 1 (*strongly disagree*) to 5 (*strongly agree*), with the statements. The items were averaged to create a total score of the six dimensions. For the self-report, Cronbach's alpha was .84 for Honesty-Humility, .89 for Emotionality, .88 for Extraversion, .88 for Agreeableness, .85 for Conscientiousness, and .88 for Openness to Experience. For the other-report, Cronbach's alpha was .85 for Honesty-Humility, .85 for Emotionality, .78 for Extraversion, .85 for Agreeableness, .83 for Conscientiousness, and .81 for Openness to Experience.

Work outcomes

We used the 21-item measure of Williams and Anderson (1991) to assess task performance (seven items) and OCB (14 items). Participants were asked for their agreement (1 = *strongly disagree*;

5 = *strongly agree*) with the statements. An example item is “The employee adequately completes assigned duties.” The items were averaged to create a total score of the two work outcomes. Cronbach’s alpha was .83 for task performance and .80 for OCB. Considering ratings of task performance and OCB are strongly influenced by liking (e.g., Sutton, Baldwin, Wood, & Hoffman, 2013), we also asked the supervisors to what extent they liked each of their employees on a scale from 1 (*not at all*) to 5 (*very much*) and controlled for this variable in the regression analyses.

Study 3 results

A CFA confirmed that a six-factor model fit the data reasonably well, $\chi^2(215) = 237.95$, $p > .05$, CFI = .81, RMSEA = .03, SRMR = .08.⁴ Item loadings varied between $-.01$ and $.84$; 13 items had an item loading greater than $.30$. The six-factor model fit the data significantly better than a one-factor model, $\Delta\chi^2(15) = 59.35$, $p < .01$; $\chi^2(230) = 297.30$, $p < .01$; CFI = .45, RMSEA = .05, SRMR = .09.

Table 7 presents the means, standard deviations, and intercorrelations of the HEXACO dimensions as measured with the SJT, the self-reported HEXACO dimensions, the other-reported HEXACO dimensions, task performance, and OCB. Based on the correlations with the self-reported HEXACO dimensions, results showed support for the construct-related validity of the SJT. The convergent correlations ranged from $.30$ ($p < .01$) for Agreeableness to $.64$ ($p < .01$) for Openness to Experience. The discriminant correlations ranged from $-.16$ ($p > .05$) between Emotionality as measured with the SJT and Agreeableness as measured with the HEXACO-PI-R to $.28$ ($p < .01$) between Agreeableness as measured with the SJT and Conscientiousness as measured with the HEXACO-PI-R. Table 2 (Study 3) shows that for all constructs the convergent correlation was higher than the average discriminant correlation. Steiger’s z showed a significant difference (all $ps < .01$) between the two correlations for all constructs, except for Honesty-Humility and Agreeableness. The convergent correlation for Agreeableness was marginally higher than the discriminant correlation ($z = 1.63$, $p > .05$).

Based on the correlations with the other-reported HEXACO dimensions, results showed support for the construct-related validity of four of the SJT scales—Honesty-Humility, Emotionality, Extraversion, and Openness to Experience. For these scales, the convergent correlations ranged from $.22$ ($p < .01$) for Honesty-Humility to $.39$ ($p < .01$) for both Emotionality and Openness to Experience. The discriminant correlations ranged from $-.04$ ($p = .68$) between Openness to Experience as measured with the SJT and Emotionality as measured with the HEXACO-PI-R to $.27$ ($p < .01$) between Openness to Experience as measured with the SJT and Honesty-Humility as measured with the HEXACO-PI-R. For the Agreeableness scale of the SJT, the convergent correlation was nonsignificant and lower than the discriminant correlations. For the Conscientiousness scale of the SJT, the convergent correlation was significant ($r = .26$, $p < .01$) but similar to the discriminant correlation with Extraversion as measured with the HEXACO-PI-R ($r = .27$, $p < .01$). Table 2 shows that for Emotionality ($z = 2.07$, $p < .01$) and Extraversion ($z = 2.17$, $p < .01$), the convergent correlation was significantly higher than the average discriminant correlation. The convergent correlation for Openness to Experience was marginally higher than the discriminant correlation ($z = 1.85$, $p > .05$, respectively).

In addition, we used a CFA to evaluate the convergent and discriminant validities. A starting model, specifying correlated trait factors but no method factors, yielded a relatively poor fit to the data, $\chi^2(132) = 263.75$, $p < .01$, CFI = .68, RMSEA = .10, SRMR = .11. The CTCU model yielded a better fit to the data, $\Delta\chi^2(45) = 119.12$, $p < .01$; $\chi^2(87) = 144.63$, $p < .01$; CFI = .86, RMSEA = .08, SRMR = .09. The standardized parameter estimates of the CTCU model are presented in Table 8. All trait factor loadings were substantial and significant, indicating that the three measures generally agreed in their estimation of the personality trait constructs. Only two intercorrelations among the latent HEXACO factors were significant, indicating adequate discriminant validity. As indicated by the uniqueness correlations, common method variance was lower in the SJT (M absolute $r = .10$) than in the self-report HEXACO measure (M absolute $r = .24$) and the other-report HEXACO measure (M absolute $r = .17$).

Table 7. Means, standard deviations, and intercorrelations for all variables in Study 3

	N	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. SJT-H	110	3.63	0.25	.20																			
2. SJT-E	110	1.95	0.39	.09	.27																		
3. SJT-X	110	2.87	0.41	.17	.05	.39																	
4. SJT-A	110	2.30	0.42	.14	-.10	-.03	.44																
5. SJT-C	110	2.95	0.44	.07	.19*	.05	.20*	.20															
6. SJT-O	110	2.84	0.50	.16	.21*	.27**	.06	.14	.48														
7. SR-H	110	3.68	0.36	.31*	.08	.02	.18	.16	-.05	.84													
8. SR-E	110	2.81	0.44	.13	.58**	-.06	.08	.22*	-.02	.03	.89												
9. SR-X	110	3.72	0.37	.07	-.06	.56**	-.06	-.04	.07	-.04	-.37**	.88											
10. SR-A	110	3.57	0.37	.16	-.16	.06	.30**	-.11	-.14	.25**	-.08	.10	.88										
11. SR-C	110	3.02	0.46	.06	.07	.00	.28**	.42**	-.07	.33**	.10	.01	.12	.85									
12. SR-O	110	3.03	0.46	.11	.08	.26**	-.04	.09	.64**	-.19*	-.09	.22*	.01	-.15	.88								
13. OR-H	96	3.72	0.53	.22*	.03	.04	.07	-.03	.27**	.45**	-.12	-.02	.13	.06	.23	.85							
14. OR-E	96	2.80	0.55	.06	.39**	.07	.11	.12	-.04	-.01	.53**	.03	-.05	.18	-.12	-.18	.85						
15. OR-X	96	3.69	0.43	.05	.11	.36**	-.06	.27**	.21*	-.08	-.01	.46**	-.03	-.08	.28**	.00	-.03	.78					
16. OR-A	96	2.97	0.53	.11	.08	.14	-.01	-.09	.18	.09	.05	.08	.40**	-.04	.24*	.24*	-.23*	.06	.85				
17. OR-C	96	3.79	0.50	.14	.10	-.01	.15	.26**	.12	.30**	.18	-.19	.08	.45**	-.04	.41**	.16	-.09	-.09	.83			
18. OR-O	96	2.82	0.56	.08	.16	.10	-.03	.01	.39**	-.15	.10	-.12	-.01	-.22*	.44**	.06	.08	.25*	.14	.03	.81		
19. Task	103	4.10	0.44	.06	.15	-.10	.04	.05	.18	.18	.02	.00	.13	.02	.11	.31**	-.03	.21*	.07	.23*	.04	.83	
20. OCB	103	3.95	0.37	.26**	.10	.07	.04	-.07	.24*	.14	-.03	.15	.22*	-.03	.25*	.38**	-.16	.28**	.15	.25*	.20	.60**	.80

Note. The situational judgment test (SJT) HEXACO dimensions were measured on a scale from 1 to 4; the self-reported (SR) and the other-reported (OR) HEXACO dimensions, task performance (task), and organizational citizenship behavior (OCB) were measured on a scale from 1 to 5. Cronbach's alphas are presented on the diagonal between parentheses. Significant convergent correlations are presented in boldface.

* $p < .05$ (two-tailed). ** $p < .01$.

Table 8. Correlated traits–correlated uniqueness standardized parameter estimates in Study 3

	H	E	X	A	C	O
Factor loadings						
SJT	.69**	.77**	.67**	.54**	.61**	.71**
HEXACO self-report	.50**	.77**	.75**	.58**	.71**	.79**
HEXACO other-report	.35**	.60**	.62**	.44**	.54**	.58**
Latent factor correlations						
H	1.00					
E	.10	1.00				
X	.16	.06	1.00			
A	.36*	-.04	.09	1.00		
C	.15	.32*	.07	.16	1.00	
O	.27	.04	.25	.11	.02	1.00
Uniqueness correlations						
SJT						
H	1.00					
E	-.11	1.00				
X	.10	.14	1.00			
A	-.01	-.21	-.08	1.00		
C	-.06	.10	-.04	.11	1.00	
O	.19	.34*	.24	.06	.06	1.00
HEXACO self-report						
H	1.00					
E	-.10	1.00				
X	-.01	-.74**	1.00			
A	.19	-.14	.20	1.00		
C	.35*	-.30	.27	.10	1.00	
O	-.35*	-.16	.22	.22	-.19	1.00
HEXACO other-report						
H	1.00					
E	-.17	1.00				
X	.03	-.17	1.00			
A	.20	-.38**	.03	1.00		
C	.49**	.06	-.06	-.15	1.00	
O	-.13	.17	.28*	.01	.04	1.00

Note. SJT = situational judgment test.

* $p < .05$. ** $p < .01$.

Two of the SJT scales significantly correlated with OCB, that is, Honesty-Humility ($r = .26$, $p < .01$) and Openness to Experience ($r = .24$, $p < .05$). Of the self-reported HEXACO dimensions, Agreeableness ($r = .22$, $p < .05$) and Openness to Experience ($r = .25$, $p < .05$) significantly correlated with OCB. Of the other-reported HEXACO dimensions, Honesty-Humility ($r = .38$, $p < .01$ and $r = .33$, $p < .01$), Extraversion ($r = .21$, $p < .05$ and $r = .28$, $p < .01$), and Openness to Experience ($r = .23$, $p < .05$ and $r = .25$, $p < .05$) significantly correlated with both task performance and OCB. Other-reports of the HEXACO dimensions explained the largest part of the variance in both task performance ($R^2 = .16$), $F(6, 82) = 2.58$, $p < .05$, and OCB ($R^2 = .26$), $F(6, 82) = 4.90$, $p < .01$, compared to the self-reports ($R^2 = .07$), $F(6, 96) = 1.13$, $p > .05$, and ($R^2 = .15$), $F(6, 96) = 2.71$, $p < .05$, respectively, and the SJT ($R^2 = .07$), $F(6, 96) = 1.20$, $p > .05$, and $R^2 = .12$, $F(6, 96) = 2.27$, $p < .05$, respectively).

To compare the criterion-related validity of the SJT to the criterion-related validity of the self-reported HEXACO measure, we conducted a series of hierarchical regression analyses with liking in Step 1, the self-reported HEXACO dimensions in Step 2, and the SJT scales in Step 3 (see Tables 9 and 10). For task performance, we found a significant negative beta weight for Extraversion ($\beta = -.26$, $t = -2.18$, $p < .05$) in Step 3. For OCB, both the Honesty-Humility scale ($\beta = .19$, $t = 2.00$, $p < .05$) and the Agreeableness scale of the HEXACO ($\beta = .20$, $t = 2.25$, $p < .05$) explained a significant part of the variance in Step 2. However, in Step 3, we only found a significant beta weight for the Honesty-Humility scale of the SJT ($\beta = .19$, $t = 2.03$, $p < .05$). A bootstrapping procedure with 10,000 bootstrapped resamples (Preacher & Hayes, 2008) revealed that the relation

Table 9. Results of hierarchical regression analyses with managers' ratings of task performance as the outcome

	Step 1				Step 2				Step3			
	<i>B</i>	<i>SE (B)</i>	β	<i>t</i>	<i>B</i>	<i>SE (B)</i>	β	<i>t</i>	<i>B</i>	<i>SE (B)</i>	β	<i>t</i>
Step 1												
Liking	0.32	0.08	.37	4.01**	0.34	0.08	.39	4.10**	0.36	0.08	.41	4.33**
Step 2												
Honesty-Humility					0.25	0.12	.20	2.03**	0.22	0.13	.18	1.72
Emotionality					−0.02	0.10	−.02	−0.22	−0.08	0.13	−.08	−0.61
Extraversion					−0.07	0.12	−.06	−0.56	0.10	0.15	.08	0.65
Agreeableness					0.13	0.11	.12	1.21	0.18	0.12	.16	1.58
Conscientiousness					0.02	0.12	.02	0.19	0.04	0.13	.00	0.04
Openness to Experience					0.07	0.10	.08	0.76	−0.06	0.13	−.06	−0.45
Step 3												
SJT-H									−0.04	0.18	−0.02	−0.21
SJT-E									0.18	0.14	0.16	1.28
SJT-X									−0.28	0.13	−0.26	−2.18*
SJT-A									0.04	0.11	0.04	0.37
SJT-C									0.01	0.11	0.01	0.07
SJT-O									0.19	0.12	0.22	1.71
R^2			.14				.21				.29	
$F (df)$			16.08** (1,101)				3.53** (7,95)				2.76** (13,89)	
ΔR^2							.07				.08	
$\Delta F (df)$							1.38 (6,95)				1.68 (6,89)	

Note. $N = 103$. SJT = situational judgment test.

* $p < .05$. ** $p < .01$.

Table 10. Results of hierarchical regression analyses with managers' ratings of organizational citizenship behavior as the outcome

	Step 1				Step 2				Step 3			
	<i>B</i>	<i>SE (B)</i>	β	<i>t</i>	<i>B</i>	<i>SE (B)</i>	β	<i>t</i>	<i>B</i>	<i>SE (B)</i>	β	<i>t</i>
Step 1												
Liking	0.31	0.07	0.43	4.76**	0.30	0.07	.42	4.62**	0.33	0.06	.45	5.06**
Step 2												
Honesty-Humility					0.19	0.10	.19	2.00*	0.11	0.10	.11	1.12
Emotionality					−0.01	0.08	−.01	−0.10	−0.05	0.10	−.07	−0.55
Extraversion					0.06	0.10	.06	0.66	0.14	0.12	.14	1.21
Agreeableness					0.19	0.08	.20	2.25*	0.17	0.09	.19	1.94
Conscientiousness					−0.02	0.09	−.02	−0.24	0.02	0.10	.02	0.21
Openness to Experience					0.14	0.08	.17	1.84	0.03	0.10	.04	0.34
Step 3												
SJT-H									0.28	0.14	.19	2.03*
SJT-E									0.10	0.11	.10	0.92
SJT-X									−0.16	0.10	−.18	−1.59
SJT-A									0.05	0.09	.05	0.55
SJT-C									−0.10	0.09	−.11	−1.14
SJT-O									0.13	0.09	.18	1.51
R^2			.18				.30				.38	
$F (df)$			22.68** (1, 101)				5.86** (7, 95)				4.24** (13,89)	
ΔR^2							.12				.08	
$\Delta F (df)$							2.68* (6, 95)				1.94 (6, 89)	

Note. $N = 103$. SJT = situational judgment test.

* $p < .05$. ** $p < .01$.

between the Honesty-Humility scale and OCB was mediated by the Honesty-Humility scale of the SJT (indirect effect = .08, $SE = .04$), 95% confidence interval [.02, .20]. The same bootstrapping procedure did not reveal a significant indirect effect of the Honesty-Humility scale of the SJT through the self-reported HEXACO scale.

Study 3 discussion

The results of Study 3 show further evidence for the construct-related and criterion-related validity of our HEXACO-SJT. Again, we found an adequate factor structure of our HEXACO-SJT. The CTCU showed adequate results and the MTMM correlations showed a satisfactory pattern between five SJT scales and self-reported HEXACO dimensions and between three SJT scales and other-reported HEXACO dimensions. Agreeableness tends to be the trait that is the hardest to perceive by others (e.g., Connelly & Ones, 2010; Connolly, Kavanagh, & Viswesvaran, 2007), which could explain the low convergent validity for the Agreeableness scale with other-reported Agreeableness.

Although other-reported personality dimensions were found to be the best predictors of both task performance and OCB, two of the SJT scales significantly correlated with OCB, that is, Honesty-Humility and Openness to Experience. The Honesty-Humility scale of the SJT explained a significant part of the variance in OCB even after controlling for self-reported HEXACO dimensions. In addition, the Honesty-Humility scale of the SJT mediated the direct effect of self-reported Honesty-Humility on OCB. These positive effects for the Honesty-Humility scale were in line with our expectation and with previous research by Day and Silverman (1989), showing that for accountants modesty and a low need to influence and control others are highly valued traits. However, in contrast with our expectations, we did not find a positive correlation between the Conscientiousness scale of the SJT and OCB, nor between the Conscientiousness scale of the HEXACO self-report and OCB. Our sample of accountants was critical of their own levels of Conscientiousness; the scores on the self-reported HEXACO scale ($M = 3.56$, $SD = 0.37$) were significantly lower than the scores on the other-reported HEXACO scale ($M = 3.79$, $SD = 0.50$), $t(94) = -4.57$, $p < .01$, $d = -0.71$. They probably used their peers as a reference group, resulting in downward biased Conscientiousness scores.

Study 4 introduction

Studies 1–3 provided insight into the construct-related and criterion-related validity of the SJT. However, an important psychometric test property and determinant of predictive validity—test-retest reliability (McCrae, Kurtz, Yamagata, & Terracciano, 2011)—is less frequently reported in SJT research. Hence, the first goal of Study 4 was to test the test-retest reliability of the HEXACO-SJT.

Study 4 also deals with a potential limitation of the current SJT; the test might be transparent because the response options are presented in order of trait expression (i.e., either from low to high trait expression or vice versa). It could be that after a few items, participants are able to discern this pattern. Therefore, the second goal of Study 4 was to compare the test-retest reliability and factor structure of the SJT we used in Studies 1–3 (Condition 1) with a slightly altered version of the SJT in which the response options were presented in a randomized order (Condition 2). We also compare the two SJT versions in terms of content difficulty (i.e., mean scores) and perceived difficulty of the task completion process.

Study 4 method

Sample and procedure

As we again tested the factor structure of the SJT, we aimed for a minimum sample size of 105 per condition (MacCallum et al., 1996). Data were collected through Amazon Mechanical Turk. A total of 317 participants completed an online survey in exchange for \$0.75. Two weeks later, we contacted the same participants again and asked them to complete a second survey in exchange for \$1.50. We removed 12 cases because of failed attention checks (“Please click ‘disagree’ here”). Our final sample consisted of 305 participants (159 men), with a minimum age of 19 and a maximum age of 70 years old ($M = 38.87$, $SD = 11.85$). Most participants were currently employed (83.60%) and their work experience ranged between 1 and 52 years ($M = 18.22$, $SD = 11.13$). A total of 244 participants

completed the survey at both T1 and T2. The content of the two surveys were identical, participants completed the HEXACO-SJT and a short scale to assess the perceived difficulty of the test. At both occasions, participants were randomly assigned to one of two conditions: the condition in which the SJT response options were presented in the original order ($n = 149$ at T1, which reduced to $n = 121$ at T2) or the condition in which the SJT response options were presented in a randomized order ($n = 156$ at T1, which reduced to $n = 123$ at T2).

Measures

SJT

We used the same SJT as in Study 1.

Perceived difficulty

This variable was measured with four items adopted from Wiechmann and Ryan (2003). An example item is “This test was hard.” The answer scale ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). Cronbach’s alpha was .83 at T1 and at T2.

Study 4 results

Table 11 presents the means, standard deviations, Cronbach’s alphas, and test–retest coefficients for the SJT versions. For the version of the SJT that we used in Studies 1–3, we found that the test–retest coefficients varied between .55 and .74. The test–retest coefficients of the SJT with the response options in a randomized order varied between .52 and .81. Changing the order of the response options at T2, led to minimal changes in test–retest coefficients: The average test–retest coefficient ranged from .60 to .65 across conditions.

A CFA confirmed that for the original SJT version both at T1 and T2, a six-factor model fit the data reasonably well and significantly better than a one-factor model: original version T1, $\chi^2(237) = 322.27$, $p < .01$, CFI = .80, RMSEA = .05, SRMR = .08, $\Delta\chi^2(15) = 92.48$, $p < .01$; original version T2, $\chi^2(237) = 324.85$, $p < .01$, CFI = .80, RMSEA = .06, SRMR = .09, $\Delta\chi^2(15) = 161.07$, $p < .01$. At T1, item loadings varied between $-.36$ and $.73$ and 19 items had an item loading greater than .30. At T2, item loadings varied between $.12$ and $.95$, and 21 items had an item loading greater than .30. For the SJT version with the response options in a randomized order, a six-factor model fit the data reasonably well and significantly better than a one-factor model at T2, $\chi^2(237) = 299.21$, $p < .01$, CFI = .79, RMSEA = .05, SRMR = .08, $\Delta\chi^2(15) = 74.27$, $p < .01$. Item loadings varied between $-.02$ and $.72$, and 18 items had an item loading greater than .30. However, at T1 the six-factor model was unidentified.

To compare the content difficulty of the two SJT versions, we examined mean score differences. Independent sample t tests revealed that at T1 and compared to the original SJT, participants scored higher on the Honesty-Humility dimension ($t = -8.80$, $p < .01$, $d = 1.02$) and the Agreeableness dimension ($t = -2.79$, $p < .01$, $d = 0.34$) on the SJT with the response options in a randomized order. At T2, participants scored higher on the Emotionality dimension ($t = -2.70$, $p < .01$, $d = 0.35$) on the SJT with the response options in a randomized order as compared to the original SJT. Next, we examined mean differences on perceived difficulty. Independent sample t tests revealed no significant differences at T1 ($t = -0.38$, $p > .05$, $d = -0.04$) or at T2 ($t = 1.54$, $p > .05$, $d = 0.20$).

Study 4 discussion

The results of Study 4 show that the test–retest reliability of the HEXACO-SJT varies between .55 and .74. These test–retest coefficients are comparable to the average test–retest coefficient of .61 that Campion, Ployhart, and MacKenzie (2014) reported in their SJT review. Furthermore, Study 4 revealed that changing the order of the response options did not improve the psychometric

Table 11. Means, standard deviations, and test–retest coefficients in Study 4

	T1				T2				Test–Retest Coefficient Same order	Test–Retest Coefficient Other order
	N	M	SD	α	N	M	SD	α		
Original order										
SJT-H	149	2.55	0.56	.27	121	3.16	0.58	.54	.55	.51
SJT-E	149	2.08	0.49	.38	121	2.05	0.48	.47	.60	.62
SJT-X	149	2.49	0.55	.43	121	2.48	0.61	.59	.68	.49
SJT-A	148	2.36	0.55	.32	121	2.45	0.59	.51	.61	.63
SJT-C	148	2.90	0.62	.60	121	2.95	0.58	.55	.60	.58
SJT-O	148	3.05	0.64	.58	121	3.03	0.65	.65	.74	.71
							Average		.64	.60
Random order										
SJT-H	150	3.13	0.59	.42	121	3.18	0.60	.48	.81	.74
SJT-E	150	2.19	0.57	.40	121	2.24	0.64	.46	.55	.54
SJT-X	150	2.56	0.57	.26	121	2.51	0.65	.45	.67	.64
SJT-A	150	2.54	0.54	.17	121	2.41	0.60	.50	.52	.41
SJT-C	150	2.99	0.54	.40	120	2.99	0.54	.34	.63	.66
SJT-O	150	2.87	0.64	.50	120	2.90	0.65	.59	.64	.68
							Average		.65	.62

Note. The situational judgment test (SJT) HEXACO dimensions were measured on a scale from 1 to 4. To calculate the average test–retest coefficient, we used the Fisher r -to- z transformation before averaging the correlations, and then changed the mean z value back to r using the same transformation.

properties of the test in terms of factor structure or test–retest reliability, nor did it change the content difficulty and perceived difficulty. In fact, compared to the original SJT, participants even scored higher on the Honesty–Humility dimension and the Agreeableness dimension on the SJT with the response options in a randomized order. It could be that randomizing the response options affected participants responses because they were forced to read each of the response options more carefully than in the original SJT.

Overall discussion

In four studies, among actual applicants, employees, and Amazon Mechanical Turk workers, we examined the construct-related validity, the criterion-related validity, and the test–retest reliability of our newly developed HEXACO-SJT. We believe our study has several theoretical and practical implications, which we outline next.

First, our research shows that it is possible to develop an SJT to measure multiple specific constructs. Although there have been other attempts of developing construct-based SJTs (e.g., Becker, 2005; Bledow & Frese, 2009; De Meijer et al., 2010), previous construct-based SJTs have focused on single constructs (e.g., integrity or personal initiative). In Studies 1–3, we show that the HEXACO-SJT measures six distinguishable constructs that show intercorrelations and factor loadings comparable to those within the self-report and other-report HEXACO measure. The convergent validities of the SJT were similar, and often higher, than those found between personality scales in general (Pace & Brannick, 2010). We therefore believe that a construct-driven SJT offers a promising alternative technique to measure personality.

A feature that distinguishes the present SJT from those in previous studies is its scoring method. In the present study, the response options were developed such that they reflected different levels of the trait ranging from 1 (the lowest level of the trait) to 4 (the highest level of the trait). Participants' SJT scores were calculated by averaging the trait scores across items. We compared the SJT scores derived from this relatively simple method to the results of SJT scores derived from a more elaborate expert-based scoring method and found similar results in terms of reliability and validity. Hence, this alternative approach, proposed by Lievens (2017), seems to be an effective and efficient method of developing and scoring personality-SJTs. Furthermore, for most SJTs and personality measures it has been hard to find a good model fit (e.g., Hopwood & Donnellan, 2010; McDaniel & Whetzel, 2005), especially in applicant settings

(e.g., MacKenzie et al., 2009; Schmit & Ryan, 1993). Hence, the satisfactory CFA results provide further support for the construct-related validity of our HEXACO-SJT.

Second, Study 3 shows that the criterion-related validity of the HEXACO-SJT is comparable to the criterion-related validity of traditional self-report measures. Two of the SJT scales (Honesty-Humility and Openness to Experience) significantly correlated with managers' ratings of OCB. In line with previous studies (e.g., Connelly & Ones, 2010; Oh et al., 2011), other-reported personality appeared to be a stronger predictor of job behavior than self-reported personality. However, other-ratings of personality are not always feasible to use for selection purposes (Mount, Barrick, & Strauss, 1994). Hence, for practical (selection) purposes, it is probably more adequate to compare the criterion-related validity of the SJT to that of the traditional self-report measure. Of all self-report scales, the Honesty-Humility scale of the SJT was the best predictor of OCB. Of interest, the Honesty-Humility scale of the SJT mediated the effect of self-reported Honesty-Humility on OCB. This finding may indicate that contextualized measures such as SJTs are more closely aligned with work contexts than traditional personality measures and therefore have the potential to be stronger and more direct predictors of job behaviors.

Limitations and suggestions for future research

Three limitations should be noted. First, the low Cronbach's alphas of the HEXACO SJT in our four studies are an important limitation of our research. As Cronbach's alpha is regarded an important test criterion—especially in high-stakes selection settings—we recommend future studies to increase the alphas by developing additional items to measure each of the HEXACO dimensions. Test-retest coefficients in Study 4 showed promising reliability estimates. For the version of the SJT that we used throughout our studies, we found that the test-retest coefficients (with a time interval of 2 weeks) varied between .55 and .74. These test-retest coefficients are comparable to test-retest coefficients that have been reported in previous SJT studies (e.g., Campion et al., 2014). Second, a concurrent design was used to examine the criterion-related validity of the SJT. Tett and Christiansen (2007) showed that faking under true hiring conditions attenuates the validity of personality measures. Thus, for the SJT to be used as a selection tool, it is important to examine its validity for predicting future job behaviors among actual applicants. Third, another potential concern related to our HEXACO-SJT, and construct-driven SJTs in general, is that the test might become more transparent because the item options reflect different levels of the traits. Although the Honesty-Humility scale of the SJT showed promising results regarding its criterion-related validity, the means for this scale were especially high (e.g., $M = 3.63$ in Study 3 and $M = 3.55$ in Study 2). Randomizing the order of the response options did not make the scale less transparent. In fact, participants even scored higher on the Honesty-Humility scale of the SJT with the response options in a randomized order than on the Honesty-Humility scale of the original SJT. Therefore, more research is needed on the effects of item transparency, faking, and retest and coaching effects of construct-driven SJTs, especially in high-stakes settings.

Despite these limitations, we believe the current HEXACO-SJT has the potential to advance both SJT and personality research. A construct-based SJT development approach, as we used in the present study, provides better insight into what the SJT is measuring and why it is related to criteria of interest than a work-sample-based SJT development approach. As a consequence, construct-based SJTs offer the opportunity for theory testing, establishing the job relevancy of the test, generalizing its validity evidence to other jobs and industries, and providing more accurate feedback to candidates (Arthur & Villado, 2008; Christian et al., 2010; Guenole et al., 2017; Lievens, 2017).

Notes

1. Note, however, that this research may confound method (e.g., response format) with construct (e.g., personality, work attitudes, job knowledge) effects.
2. Because it is unlikely that in every SJT item the four response options would cover the entire trait continuum, we also used a more elaborate scoring approach, that is, by assigning scores based on trait loadings as rated by

the three authors of this article. The ICCs (3,3) of these ratings were substantial: .96 for Honesty-Humility, .96 for Emotionality, .99 for Extraversion, .98 for Agreeableness, .95 for Conscientiousness, and .99 for Openness to Experience. However, this alternative scoring approach showed highly similar results to the ones reported in terms of reliability and validity in all studies. For reasons of parsimony, we report only the results for the original scoring approach as suggested by Lievens (2017).

3. Apart from the six domain scales, these two versions (a full and half-length version) contain two interstitial facets, Altruism and Proactivity. Results on these two facets can be obtained from the first author.
4. Note that we had to remove one item of the Honesty-Humility scale for the CFA analyses, because this item did not show any variance.

Acknowledgments

We thank Danny van der Horst, Myckel Cremers, Barbara Rosa, Tirsia van der Linden, Barend Koch, Diederik Gosewehr, and Antonet Lajqi for their help with data collection.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Anglim, J., Morse, G., De Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing job applicants to non-applicants using an item-level bifactor model on the HEXACO Personality Inventory. *European Journal of Personality*, 31, 669–684. doi:10.1002/per.2120
- Arthur, J. W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435–442. doi:10.1037/0021-9010.93.2.435
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11, 150–166. doi:10.1177/1088868306294907
- Ashton, M. C., & Lee, K. (2008). The prediction of Honesty–Humility-related criteria by the HEXACO and Five-Factor Models of personality. *Journal of Research in Personality*, 42, 1216–1228. doi:10.1016/j.jrp.2008.03.006
- Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, 91, 340–345. doi:10.1080/00223890902935878
- Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, 18, 139–152. doi:10.1177/1088868314523838
- Ashton, M. C., Lee, K., De Vries, R. E., Perugini, M., Gnisci, A., & Sergi, I. (2006). The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English. *Journal of Research in Personality*, 40, 851–875. doi:10.1016/j.jrp.2005.06.003
- Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, 87, 707–721. doi:10.1037/0022-3514.87.5.707
- Becker, T. E. (2005). Development and validation of a situational judgment test of employee integrity. *International Journal of Selection and Assessment*, 13, 225–232. doi:10.1111/j.1468-2389.2005.00319.x
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317–335. doi:10.1111/j.1468-2389.2006.00354.x
- Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, 62, 229–258. doi:10.1111/j.1744-6570.2009.01137.x
- Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10, 99–109. doi:10.1207/s15327043hup1002_3
- Campion, M. C., Ployhart, R. E., & MacKenzie, W. I., Jr. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance*, 27, 283–310. doi:10.1080/08959285.2014.929693
- Chiaburu, D. S., Oh, I. S., Berry, C. M., Li, N., & Gardner, R. G. (2011). The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology*, 96, 1140–1166. doi:10.1037/a0024004
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83–117. doi:10.1111/j.1744-6570.2009.01163.x
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410–417. doi:10.1037/0021-9010.86.3.410
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi:10.1037/0033-2909.112.1.155

- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122. doi:10.1037/a0021212
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15, 110–117. doi:10.1111/j.1468-2389.2007.00371.x
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4, 5–13. doi:10.1037/1040-3590.4.1.5
- Day, D. V., & Silverman, S. B. (1989). Personality and job performance: Evidence of incremental validity. *Personnel Psychology*, 42, 25–36. doi:10.1111/j.1744-6570.1989.tb01549.x
- De Cuyper, K., De Houwer, J., Vansteelandt, K., Perugini, M., Pieters, G., Claes, L., & Hermans, D. (2017). Using indirect measurement tasks to assess the self-concept of personality: A systematic review and meta-analyses. *European Journal of Personality*, 31, 8–41. doi:10.1002/per.2092
- De Meijer, L. A., Born, M. P., Van Zielst, J., & Van der Molen, H. T. (2010). Construct-driven development of a video-based situational judgment test for integrity. *European Psychologist*, 15, 229–236. doi:10.1027/1016-9040/a000027
- De Raad, B., Barelds, D. P., Timmerman, M. E., De Roover, K., Mlačić, B., & Church, A. T. (2014). Towards a pan-cultural personality structure: Input from 11 psycholexical studies. *European Journal of Personality*, 28, 497–510. doi:10.1002/per.1953
- De Vries, R. E. (2012). Personality predictors of leadership styles and the self-Other agreement problem. *The Leadership Quarterly*, 23, 809–821. doi:10.1016/j.leaqua.2012.03.002
- De Vries, R. E. (2013). The 24-item brief HEXACO inventory (BHI). *Journal of Research in Personality*, 47, 871–880. doi:10.1016/j.jrp.2013.09.003
- De Vries, R. E., Ashton, M. C., & Lee, K. (2009). De zes belangrijkste persoonlijkheidsdimensies en de HEXACO persoonlijkheidsvragenlijst [The six most important personality dimensions and the HEXACO personality questionnaire]. *Gedrag & Organisatie*, 3, 232–274. Retrieved May 2, 2016, from https://www.gedragenorganisatie.nl/inhoud/tijdschrift_artikel/GO-22-3-4
- De Vries, R. E., Lee, K., & Ashton, M. C. (2008). The Dutch HEXACO Personality Inventory: Psychometric properties, self-Other agreement, and relations with psychopathy among low and high acquaintanceship dyads. *Journal of Personality Assessment*, 90, 142–151. doi:10.1080/00223890701845195
- De Vries, R. E., Tybur, J. M., Pollet, T. V., & Van Vugt, M. (2016). Evolution, situational affordances, and the HEXACO model of personality. *Evolution & Human Behavior*, 37, 407–421. doi:10.1016/j.evolhumbehav.2016.04.001
- Eschleman, K. J., & Burns, G. (2012). The utility of general and school-specific personality traits and an examination of their relationship over time. *Learning and Individual Differences*, 22, 537–543. doi:10.1016/j.lindif.2012.03.007
- Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59, 1216–1229. doi:10.1037/0022-3514.59.6.1216
- Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, 17, 1–19. doi:10.1080/15305058.2017.1297817
- Holtrop, D. J., Born, M. P., De Vries, A., & De Vries, R. E. (2014). A matter of context: A comparison of two types of contextualized personality measures. *Personality and Individual Differences*, 68, 234–240. doi:10.1016/j.paid.2014.04.029
- Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of SJTs: Faking, coaching and retesting issues. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests. Theory, measurement, and application* (pp. 205–232). Mahwah, NJ: Erlbaum.
- Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14, 332–346. doi:10.1177/1088868310361240
- Kasten, N., & Freund, P. A. (2015). A meta-analytical multilevel reliability generalization of situational judgment tests (SJTs). *European Journal of Psychological Assessment*, 32(230–240). doi:10.1027/1015-5759/a000250
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165–172. doi:10.1037/0033-2909.112.1.165
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39, 329–358. doi:10.1207/s15327906mbr3902_8
- Lee, K., & Ashton, M. C. (2013). Prediction of self-and observer report scores on HEXACO-60 and NEO-FFI scales. *Journal of Research in Personality*, 47, 668–675. doi:10.1016/j.jrp.2013.06.002
- Lievens, F. (2017). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing*, 17, 1–7. doi:10.1080/15305058.2017.1309857
- Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology*, 91, 247–258. doi:10.1037/0021-9010.91.2.247
- Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268–279. doi:10.1037/0021-9010.93.2.268
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86–92. doi:10.1027/1614-2241.1.3.86

- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149. doi:[10.1037/1082-989X.1.2.130](https://doi.org/10.1037/1082-989X.1.2.130)
- MacKenzie, W. I., Jr, Ployhart, R. E., Weekley, J. A., & Ehlers, C. (2009). Contextual effects on SJT responses: An examination of construct validity and mean differences across applicant and incumbent contexts. *Human Performance*, 23, 1–21. doi:[10.1080/08959280903400143](https://doi.org/10.1080/08959280903400143)
- Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks, CA: Sage.
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal consistency, retest reliability, and their implications for personality scale validity. *Personality and Social Psychology Review*, 15, 28–50. doi:[10.1177/1088868310366253](https://doi.org/10.1177/1088868310366253)
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91. doi:[10.1111/j.1744-6570.2007.00065.x](https://doi.org/10.1111/j.1744-6570.2007.00065.x)
- McDaniel, M. A., List, S. K., & Kepes, S. (2016). The ‘hot mess’ of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology*, 9, 47–51. doi:[10.1017/iop.2015.115](https://doi.org/10.1017/iop.2015.115)
- McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, 33, 515–525. doi:[10.1016/j.intell.2005.02.001](https://doi.org/10.1016/j.intell.2005.02.001)
- Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 80, 252–283. doi:[10.1037/h0035002](https://doi.org/10.1037/h0035002)
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029–1049. doi:[10.1111/j.1744-6570.2007.00100.x](https://doi.org/10.1111/j.1744-6570.2007.00100.x)
- Motowidlo, S. J., Ghosh, K., Mendoza, A. M., Buchanan, A. E., & Lerma, M. N. (2016). A context-independent situational judgment test to measure prosocial implicit trait policy. *Human Performance*, 29, 331–346. doi:[10.1080/08959285.2016.1165227](https://doi.org/10.1080/08959285.2016.1165227)
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91, 749–761. doi:[10.1037/0021-9010.91.4.749](https://doi.org/10.1037/0021-9010.91.4.749)
- Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, 79, 272–280. doi: [10.1037/0021-9010.79.2.272](https://doi.org/10.1037/0021-9010.79.2.272)
- Mussel, P., Gatzka, T., & Hewig, J. (2016). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment*. Advance online publication. doi:[10.1027/1015-5759/a000346](https://doi.org/10.1027/1015-5759/a000346)
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*, 66, 574–583. doi:[10.1037/h0040291](https://doi.org/10.1037/h0040291)
- Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, 96, 762–773. doi:[10.1037/a0021832](https://doi.org/10.1037/a0021832)
- Oostrom, J. K., Born, M. P., Serlie, A. W., & Van der Molen, H. T. (2012). Implicit trait policies in multimedia situational judgment tests of leadership skills: Can they predict leadership behavior? *Human Performance*, 25, 335–353. doi:[10.1080/08959285.2012.703732](https://doi.org/10.1080/08959285.2012.703732)
- Oostrom, J. K., De Soete, B., & Lievens, F. (2015). Situational judgment testing: A review and some new developments. In I. Nikolaou & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 172–189). Sussex, UK: Psychology Press.
- Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the “same” construct? A meta-analytic investigation. *Personality and Individual Differences*, 49, 669–676. doi:[10.1016/j.paid.2010.06.014](https://doi.org/10.1016/j.paid.2010.06.014)
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). London, UK: Guilford.
- Pomerance, M. H., & Converse, P. D. (2014). Investigating context specificity, self-schema characteristics, and personality test validity. *Personality and Individual Differences*, 58, 54–59. doi:[10.1016/j.paid.2013.10.005](https://doi.org/10.1016/j.paid.2013.10.005)
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879–891. doi:[10.3758/BRM.40.3.879](https://doi.org/10.3758/BRM.40.3.879)
- Ryan, A. M., Inceoglu, I., Bartram, D., Golubovich, J., Reeder, M., Derous, E., ... Yao, X. (2015). Trends in testing: Highlights of a global survey. In I. Nikolaou & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 136–153). Sussex, UK: Psychology Press. Retrieved May 22, 2018, from <https://biblio.ugent.be/publication/5947029/file/5947030>
- Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives on Psychological Science*, 9, 538–551. doi:[10.1177/1745691614543972](https://doi.org/10.1177/1745691614543972)
- Saucier, G. (2009). Recurrent personality dimensions in inclusive lexical studies: Indications for a big six structure. *Journal of Personality*, 77, 1577–1614. doi:[10.1111/j.1467-6494.2009.00593.x](https://doi.org/10.1111/j.1467-6494.2009.00593.x)
- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology*, 78, 966–974. doi:[10.1037/0021-9010.78.6.966](https://doi.org/10.1037/0021-9010.78.6.966)
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80, 607–620. doi:[10.1037/0021-9010.80.5.607](https://doi.org/10.1037/0021-9010.80.5.607)

- Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 45–65. doi:10.1146/annurev-orgpsych-031413-091255
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, 65, 445–494. doi:10.1111/j.1744-6570.2012.01250.x
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A metaanalytic investigation of the relative validity of contextualized and non contextualized personality measures. *Personnel Psychology*, 65, 445–494. doi: 10.1111/j.1744-6570.2012.01250.x
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251. doi:10.1037/0033-2909.87.2.245
- Sutton, A. W., Baldwin, S. P., Wood, L., & Hoffman, B. J. (2013). A meta-analysis of the relationship between rater liking and performance ratings. *Human Performance*, 26, 409–429. doi:10.1080/08959285.2013.836523
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88, 500–517. doi:10.1037/0021-9010.88.3.500
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60, 967–993. doi:10.1111/j.1744-6570.2007.00098.x
- Tomas, J. M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 84–98. doi:10.1080/10705519909540120
- Tupes, E. C., & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (Technical Report ASD-TR-61-97). Lackland Air Force Base, TX: U.S. Air Force.
- Weekley, J. A., & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, 18, 81–104. doi:10.1207/s15327043hup1801_4
- Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 157–182). Mahwah, NJ: Lawrence Erlbaum.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19, 188–202. doi:10.1016/j.hrmmr.2009.03.007
- Whetzel, D. L., Mc Daniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21, 291–309. doi:10.1080/08959280802137820
- Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment*, 11, 215–229. doi: 10.1111/1468-2389.00245
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17, 601–617. doi:10.1177/014920639101700305

Appendix

The HEXACO-SJT Items

Instructions:

Below you will find 24 work situations, each followed by 4 possible ways of handling the situation. Choose the answer that most closely resembles the way you would respond. Do not think too long about your responses, your first hunch is usually the best.

We kindly ask you to choose one of the response options in each situation, even if you are not completely sure of your response.

Item 1 (H1)

You accidentally scratched the car that you borrowed from the company for one day. You are aware of the rule that the one responsible for the damage has to pay for the repair costs. What would you do?

- (a) I would go to my supervisor, be honest, and tell him that I am the one who is responsible for the damage. (4)
- (b) I would not tell my supervisor and hope that he does not notice the damage. If he does notice the damage, I would admit that I am the one who is responsible. (3)
- (c) I would not tell my supervisor and hope that he does not notice the damage. If he notices the damage I would act surprised and pretend that I know nothing about it. (2)
- (d) I would tell my supervisor that the car was already damaged when I borrowed it; he would not be able to find out who caused the damage and the organizations is rich enough to cover the repair costs. (1)

Item 2 (E1)

Tomorrow you will have your performance review, and there is a lot at stake. It is late in the evening and you are feeling tired. What would you do?

- (a) I would go to bed and fall asleep right away even though I am a bit anxious. (1)
- (b) I would go to bed but would have trouble falling asleep, once I fall asleep I manage to have a good night's rest. (2)
- (c) I would go to bed but lie awake for hours and sleep poorly. (3)
- (d) I would go to bed but lie awake all night worrying about the next day. (4)

Item 3 (X1)

During a meeting, the chair asks people to share some workplace experiences. What would you do?

- (a) I would be the first to start talking and elaborate on my experiences. (4)
- (b) After one of my colleagues has said something, I would say something too. (3)
- (c) I would only say something if I am asked personally to share my experiences. (2)
- (d) I would be hoping for my colleagues to say something so that I do not have to. (1)

Item 4 (A1)

Last year one of your colleagues took credit for business you brought in. You are about to start a new collaborative project with this colleague. What would you do?

- (a) I would have long forgotten about the incident and would start this new collaborative project with an open attitude. (4)
- (b) I would still remember the incident, but a year is a long time. I would give him a second chance. (3)
- (c) I would remember the incident as if it were yesterday. I would tell him so, but still give him a second chance. (2)
- (d) I would not be able to forget about the incident and therefore decide not to pursue this collaboration. (1)

Item 5 (C1)

You have finished a proposal that has to be submitted by tomorrow. The secretary usually checks these proposals for errors. What would you do?

- (a) I would send the proposal to the secretary and I would submit it after her check. (1)
- (b) As soon as the secretary sends back the proposal I would give it another quick check and then submit it. (2)
- (c) I would check the proposal once more before I send it to the secretary. When the secretary sends back the proposal I would give it another check and then submit it. (3)
- (d) I would give it a thorough check before I send it to the secretary. When the secretary sends back the proposal I would give it another thorough check and then submit it. (4)

Item 6 (O1)

During the annual outing you are—among other things—visiting an art gallery in the city. What would you do?

- (a) I would be very bored and would hope to leave as soon as possible. (1)
- (b) I would not enjoy it very much, but some art pieces I would appreciate. (2)
- (c) I would enjoy it and would see as many art pieces as possible. (3)
- (d) I would enjoy it very much and would hope there is enough time to see every detail of every art piece. (4)

Item 7 (H2)

Last month's salary got transferred to your bank account. When you look at your bank account you notice that you have received 500 euro too much. What would you do?

- (a) This is the company's mistake for which I am not responsible and therefore I would keep the 500 euro. (1)
- (b) I would ask my colleagues if they also received too much salary and how they are thinking of solving this. If they did not receive too much salary, I will keep the money. (2)

- (c) I would call my supervisor to indicate that I received too much money in my bank account and ask him what to do. (3)
- (d) I would transfer the money back to the company's back account right away. (4)

Item 8 (E2)

You have applied for a desirable job and you will hear soon whether they will make you an offer. What would you do?

- (a) I would calmly wait until I hear something. (1)
- (b) I would be a bit nervous, but focus on other things. (2)
- (c) I would be nervous and have difficulty concentrating on other things. (3)
- (d) I would be terribly nervous and unable to think of anything else. (4)

Item 9 (X2)

You just had two weeks of holiday and it is time to get back to work. How would you feel?

- (a) I would dread having to answer questions about my holiday again and again and therefore I would try to avoid my colleagues on the first day. (1)
- (b) Although I would look forward to seeing my colleagues again, I notice that all the attention costs me energy. (2)
- (c) I would enjoy seeing everyone again and to share my holiday experiences with my colleagues. (3)
- (d) I would be full of energy and stories and share them with all my colleagues. (4)

Item 10 (A2)

A coworker of the department that you are supervising made an important phone call with a client. You had told the coworker in advance that this was not his task and that you would make the phone call yourself. The phone call with the client was unsuccessful and now you have missed out on an important deal. What would you do?

- (a) I would be terribly upset about the fact that he made the phone call behind my back and about losing the client. I would firmly address him and tell him that his actions have consequences. (1)
- (b) I would ask the coworker to come to my office and calmly tell him that I do not appreciate his actions. (2)
- (c) I would ask the coworker to come to my office. I would tell him that I forgive him and ask what went wrong during the phone call. (3)
- (d) I would explain that I understand that coworkers find it important to have these kinds of phone calls with their own clients. I decide to give a negotiations training to all coworkers. (4)

Item 11 (C2)

You have to give a presentation to your team soon, for which you are making PowerPoint slides. What would you do?

- (a) I would only spend time on the content; details regarding layout and language are not important. (1)
- (b) I would spend a lot of time on the content and less on the details regarding layout and language. (2)
- (c) I would spend a lot of time on the content, but also some time on the details regarding layout and language. (3)
- (d) I would spend a lot of time on the content, but at least as much time on the details regarding layout and language. (4)

Item 12 (O2)

The company closed a deal with a client from Norway and asks who would like to be involved on the project. That person would have to learn some things about Norway. What would you do?

- (a) I would volunteer, because I would find it interesting to learn more about the history and culture of another country. (4)
- (b) If nobody else volunteers, I would do it, because I would find it kind of interesting to learn more about the history and culture of another country. (3)
- (c) If nobody else volunteers, I would do it, even though I am not very interested in learning more about the history and culture of another country. (2)
- (d) I would not volunteer, because I am not interested in learning more about the history and culture of another country. (1)

Item 13 (H3)

Your supervisor compliments you on a report that was actually written by one of your colleagues. What would you do?

- (a) I would thank my supervisor and leave it at that. (1)
- (b) I would thank my supervisor and tell that someone else has worked hard on it as well. (2)
- (c) I would accept the compliment, but indicate that most of the report has been written by someone else. (3)
- (d) I would not accept the compliment and tell who has actually written the report. (4)

Item 14 (E3)

Your job is at risk due to a reorganization of your company. You are very worried about what might happen. What would you do?

- (a) I would keep my worries to myself, I do not feel a need to share them with anyone. (1)
- (b) After a while I would share my worries with someone, because on closer inspection I do feel a need to do so. (2)
- (c) I would share my worries if someone asks me how I am doing and there is an opportunity to do so. (3)
- (d) I would share my worries with someone as soon as possible, this is not something I would want to keep to myself. (4)

Item 15 (X3)

On your first workday at a new company, there are a lot of new colleagues with whom to get acquainted. What would you do?

- (a) I would try to get to know the workplace first, after which I would introduce myself to my colleagues. (1)
- (b) I would quickly shake hands with my colleagues after which I would get to know the workplace. (2)
- (c) I would have a chat with all the colleagues after which I would try to get to know the workplace. (3)
- (d) I would try to get to know all the colleagues right away after which I would try to get to know the workplace. (4)

Item 16 (A3)

It appears that your colleague has requested holiday leave for the same week as you. According to your supervisor one of you has to take a different week because it would be too busy at work if both of you are absent. Your colleague is not willing to change his holiday plans. What would you do?

- (a) I would not change my holiday plans either. If my colleague is not willing to help find a solution, I am not willing to either. Your supervisor has to find a way to solve it. (1)
- (b) I would not let my colleague walk over me. If my colleague has a really good reason for why he cannot change his holiday plans, I might reconsider taking a different week, but only then. (2)
- (c) I would carefully try to talk to my colleague about the matter, but if after this my colleague would still not be willing to change his holiday plans, I will adjust my plans. (3)
- (d) I would find it annoying to change my holiday plans, but I would take a different week to prevent having a discussion with my colleague. (4)

Item 17 (C3)

You are instructed to give a presentation to three board members. You have sufficient time to prepare the presentation. What would you do?

- (a) I would start preparing the presentation an hour in advance, because I perform best under pressure. (1)
- (b) I would prepare the presentation a few days in advance and then check and practice the presentation at the last minute. (2)
- (c) I would finish the presentation way in advance and carefully check the presentation. I would practice the presentation at the last minute. (3)
- (d) I would finish the presentation way in advance. I would carefully check the presentation a few times myself and then have a colleague look at it as well. Way in advance I would run through the presentation a few times so I know it by heart. (4)

Item 18 (O3)

During a team outing you can choose between different types of workshops, including “composing a song” and “making a painting”. What would you do?

- (a) I would participate in one of these workshops, because I like doing artistic things. (4)
- (b) I would participate in one of these workshops if one of my colleagues participates as well. Could be fun to try. (3)
- (c) I would rather participate in a different workshop, one in which I do not have to do anything artistic. (2).
- (d) I would not participate in these workshops anyway, I do not like such things. (1)

Item 19 (H4)

Based on the year-end figures, you will be placed into one of the following categories depending on your performance: 1) you will receive a bonus, 2) nothing changes, or 3) your department will be curtailed. Based on the figures, it is likely that you will just fall into the 3rd category. You know that other departments usually change their figures so that they will fall into the 1st category. What would you do?

- (a) I would change the figures so that I would easily fall into the 1st category. (1)
- (b) I would change the figures so that I just fall into the 1st category. (2)
- (c) I would change the figures slightly so that I would just fall into the 2nd category. (3)
- (d) I would leave the figures as they are. (4)

Item 20 (E4)

During a personal conversation, one of your team members tells you that she and her partner are having a divorce. She is very emotional. What would you do?

- (a) I would listen and feel very bad for her, but would not get emotional myself. (1)
- (b) I would listen and sympathize with her, but I would not get emotional myself. (2)
- (c) I would listen and sympathize with her and would get teary eyes. (3)
- (d) I would listen and would personally feel her pain and (almost) start to cry myself. (4)

Item 21 (X4)

You are invited to a network meeting of one of your largest clients. You know in advance that you do not know anyone there. What would you do?

- (a) I would not go, because I would not feel comfortable if I do not know anyone. (1)
- (b) I would rather not go, unless a colleague would like to join me. (2)
- (c) I would go, because it can be useful for the company to make new contacts. (3)
- (d) I would go, because I enjoy meeting new people. (4)

Item 22 (A4)

During a meeting you are asked to take minutes. However, you already took minutes during the last meeting and you usually rotate this task. What would you do?

- (a) I would say that I took minutes last week and state that someone else has to be assigned this task. (1)
- (b) I would say that I took minutes last week. I would be open to taking minutes again unless someone else volunteers, but would make clear that I will not be taking minutes again any time soon. (2)
- (c) I would take on this task. If I am asked again at the next meeting, I would indicate that it is someone else's turn to take minutes. (3)
- (d) I would take minutes as I did during the last meeting. If I am asked again at the next meeting, I would probably not say anything about it either. They must think I am a good scribe. (4)

Item 23 (C4)

You arrive at the office in the morning and it turns out to be a big mess. What would you do?

- (a) I would not be bothered by it. (1)
- (b) I would make some space where I can work. (2)

- (c) I would organize the things I need that day. (3)
- (d) Before I start working, I would tidy up the place. (4)

Item 24 (O4)

During a lunch at work, people start a discussion about the meaning of life. What would you do?

- (a) I would withdraw and not engage in the discussion. I find philosophy boring. (1)
- (b) I would try to listen to the discussion, but quickly lose interest. (2)
- (c) I would listen and be fascinated by the discussion. I find it interesting what people have to say about it. (3)
- (d) I would actively engage in the discussion. I find philosophy very interesting. (4)